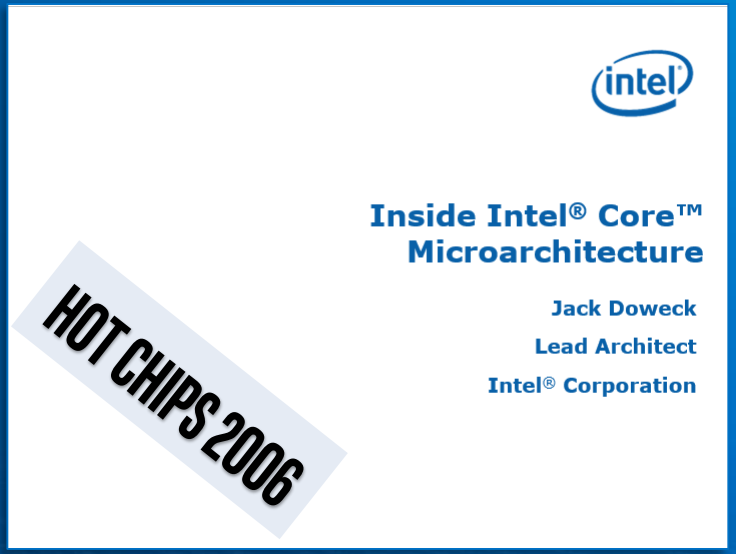
INSIDE 6TH GEN INTEL® CORE™:
NEW MICROARCHITECTURE CODE NAMED SKYLAKE

Jack Doweck, Wen-fu Kao

Intel Corporation, 2016

Authors: Ittai Anati, David Blythe, Jack Doweck, Hong Jiang, Wen-fu Kao, Julius Mandelblat, Lihu Rappoport, Efraim Rotem, Ahmad Yasin

SOME ARCHEOLOGY

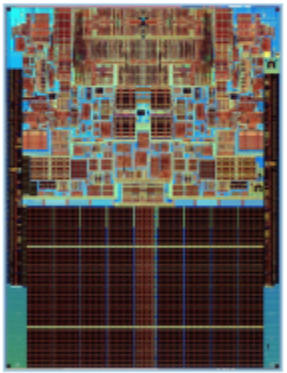


HOT CHIPS 2006

What is Intel® Core™ Microarchitecture?

The Intel® Core™ Microarchitecture is a new foundation for Intel® architecture-based desktop, mobile, and mainstream server multi-core processors

Designed for efficiency and optimized performance across a range of market segments and power envelopes



2006 Intel Core Microarchitecture based processors:

DP Server:
Dual-core Intel® Xeon® 51xx Processors
Quad-core codenamed *Clovertown*

Desktop:
Dual-core Intel® Core™ 2 Duo Processors
Quad-core codenamed *Kentsfield*

Mobile:
Dual-core Intel® Core™ 2 Duo Processors

3

Codenames are used for referring the products to indicate that the products and the tests are prerelease version and will change any time without further notice. Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.

intel

HOT CHIPS 2006

What is Intel® Core™ Microarchitecture?

The Intel® Core™ Microarchitecture is a new foundation for Intel® architecture-based desktop, mobile, and mainstream server multi-core processors

Designed for efficiency and optimized performance across a range of market segments and power envelopes

Process:
65nm

Die size:
143 mm²

Transistor count:
291 M

Execution core area:
36 mm²

Execution core transistor count:
19 M

2006 Intel Core Microarchitecture based processors:

DP Server:
Dual-core Intel® Xeon® 51xx Processors
Quad-core codenamed *Clovertown*

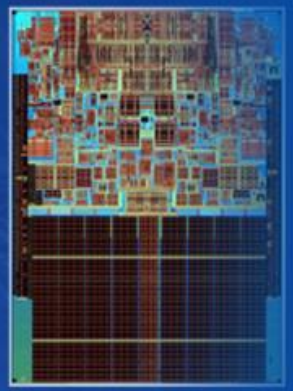
Desktop:
Dual-core Intel® Core™ 2 Duo Processors
Quad-core codenamed *Kentsfield*

Mobile:
Dual-core Intel® Core™ 2 Duo Processors

Codenames are used for referring the products to indicate that the products and the tests are prerelease version and will change any time without further notice. Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.

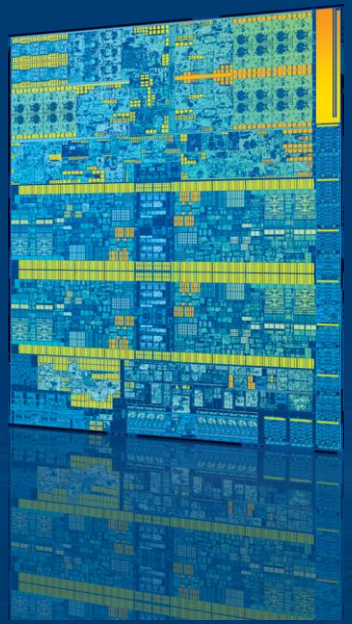
4

AN HISTORICAL PERSPECTIVE



WHAT CHANGED IN ~10 YEARS?

SKL (HOT CHIPS 2016) COMPARED TO MRM (HOT CHIPS 2006)



UP TO 10X MORE EFFICIENT

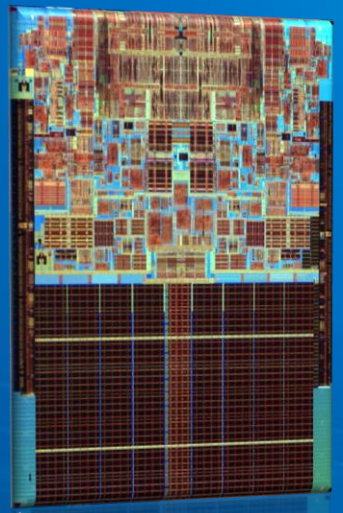
3-5X FASTER WITH 2X LOWER TDP[1]

5.0X DENSER

1.5X SMALLER

6.0X HIGHER

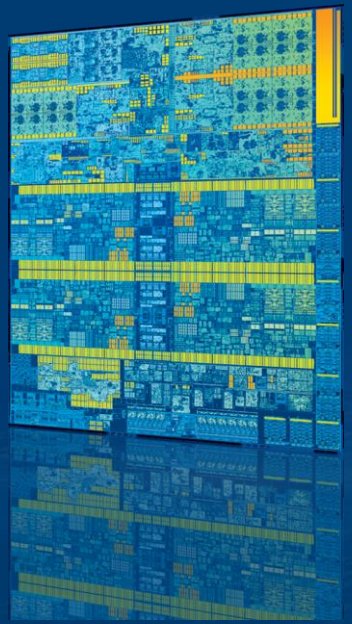
PROCESS:	65NM
DIE SIZE:	143 MM²
TRANSISTOR COUNT:	291 M



[1] Based on estimated SPECint_rate_base2006 and SPECfp_rate_base2006.
Range depends on benchmark, product and platform. Gains include advance on compiler and memory technologies.

WHAT CHANGED IN ~10 YEARS?

SKL (HOT CHIPS 2016) COMPARED TO MRM (HOT CHIPS 2006)

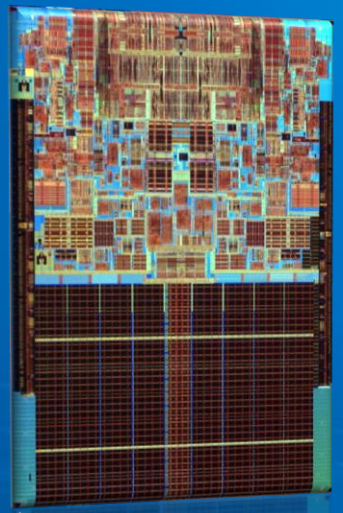


UP TO 10x MORE EFFICIENT

3-5x FASTER WITH 2x LOWER TDP[1]

1.5x SMALLER AREA

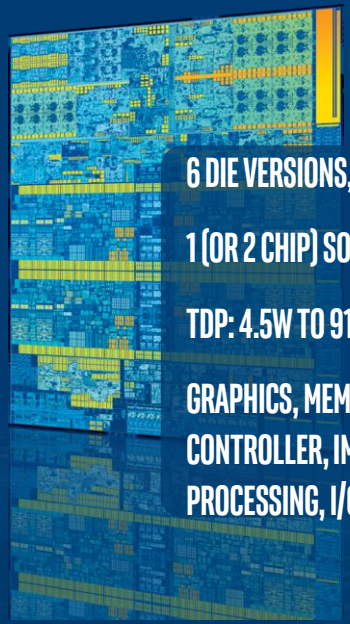
6.0x HIGHER TRANSISTOR COUNT



[1] Based on estimated SPECint_rate_base2006 and SPECfp_rate_base2006.
Range depends on benchmark, product and platform. Gains include advance on compiler and memory technologies.

WHAT CHANGED IN ~10 YEARS?

SKL (HOT CHIPS 2016)



6 DIE VERSIONS, 2 OR 4 CORES

1 (OR 2 CHIP) SOLUTION

TDP: 4.5W TO 91W

GRAPHICS, MEMORY CONTROLLER, IMAGE PROCESSING, I/OS

COMPARED TO

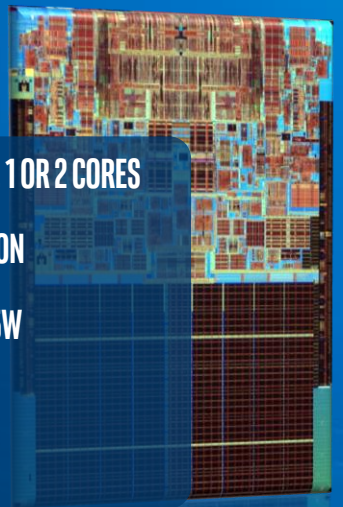
UP TO 10X MORE EFFICIENT

3-5X FASTER WITH 2X LOWER TDP[1]

1.5X SMALLER AREA

6.0X HIGHER TRANSISTOR COUNT

MRM (HOT CHIPS 2006)



3 DIE VERSIONS, 1 OR 2 CORES

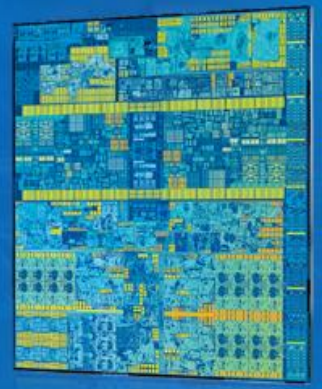
3 CHIPS SOLUTION

TDP: 5.5W TO 75W

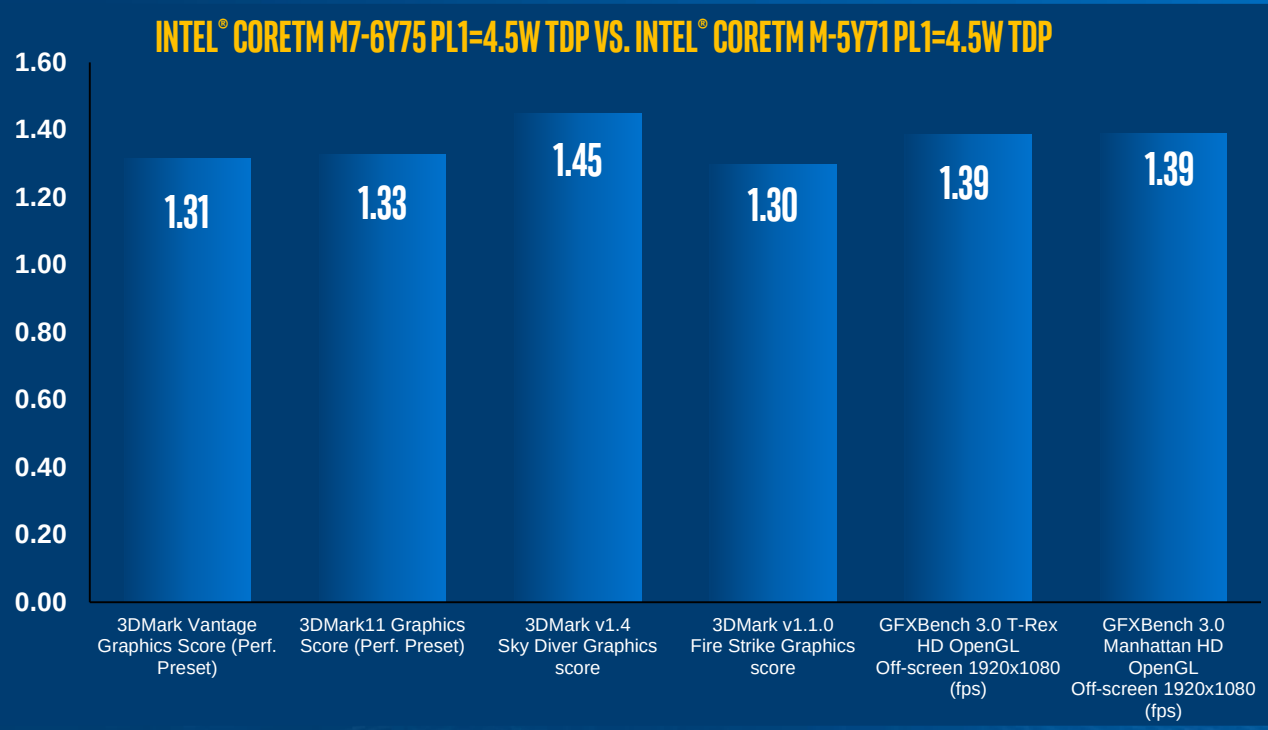
CPU ONLY

[1] Based on estimated SPECint_rate_base2006 and SPECfp_rate_base2006.
Range depends on benchmark, product and platform. Gains include advance on compiler and memory technologies.

SKYLAKE'S PERFORMANCE AND POWER



INTEL® CORE™ M7-6Y75 PROCESSOR (4.5W TDP)



In the same 4.5W chassis:

UP TO 25% MORE COMPUTE

40% BETTER 3D GAMING

Vs. prior gen

10 HOURS BATTERY LIFE[1]

Up to 10 hours of 1080p video playback

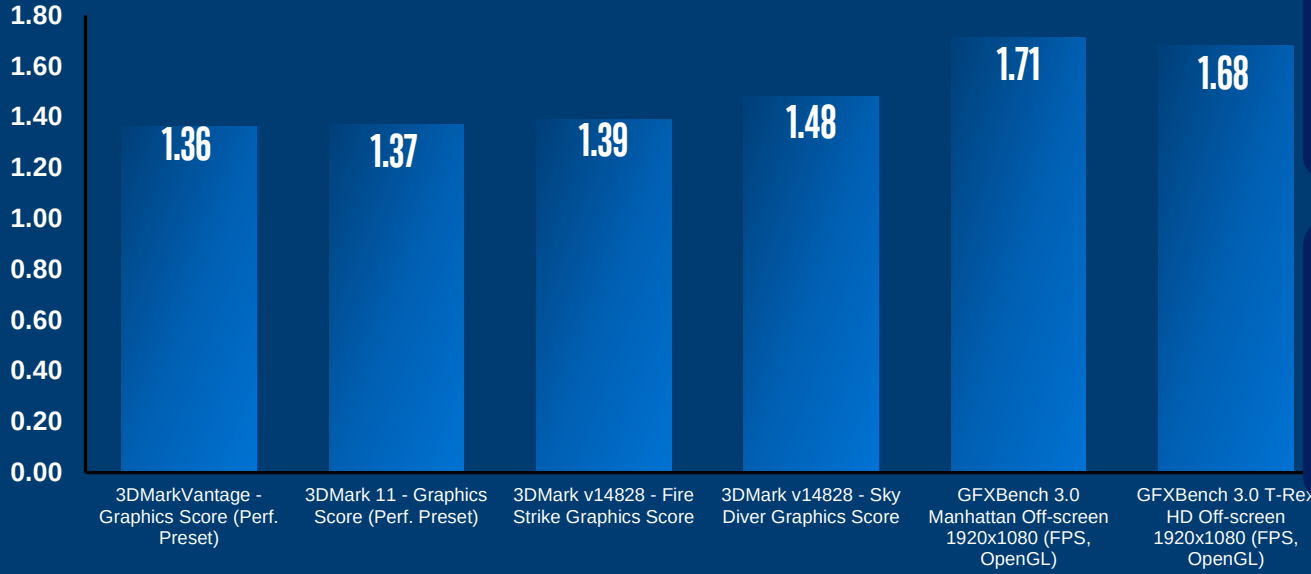
** System Configuration details:
5th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ M-5Y71processor (Turbo up to 2.9GHz/2.6GHz, PL1=4.5W TDP, 2C4T), Memory: 4GP (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows* 10.
6th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ m7-6Y75 processor (Turbo up to 3.1GHz/2.9GHz, PL1=4.5W TDP 1 PL1=4.5W TDP , 2C4T), Memory: 4GB (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows* 10.

[1] 1080p Video Battery Life Rundown projection (in hours) for Tears of Steel*, 1080p, 23.976fps, H.264, 3000kbps, rev1.mp4 >/=200nit, DPST=ON, no dimming, balanced, Wi-Fi connected. 11.6" 19x10 eDP panel and 36 WHr battery assumed.

SPEC CPU2006 estimates based on measurements on reference boards.

INTEL® CORE™ I7-6650U PROCESSOR (15W TDP)

INTEL® CORETM I7-6650U PL1=15W TDP VS. INTEL® CORETM I7-5650U PL1= 15W TDP



9 HOURS BATTERY LIFE[1]

Up to 9 hours of 1080p video playback

In a 15W chassis:

UP TO 28% MORE COMPUTE

70% BETTER 3D GAMING

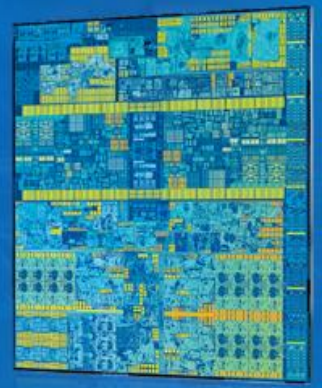
Vs. prior gen

** System Configuration details:
5th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ M-5Y71processor (Turbo up to 2.9GHz/2.6GHz, PL1=4.5W TDP, 2C4T), Memory: 4GP (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows® 10.
6th generation Intel® Core™ system: Intel Reference Platform, Intel® Core™ m7-6Y75 processor (Turbo up to 3.1GHz/2.9GHz, PL1=4.5W TDP 1 PL1=4.5W TDP , 2C4T), Memory: 4GB (2x2GB) LPDDR3-1600, HDD: Intel® Solid State Drive (Intel® SSD), Display: 1920x1080, OS: Windows® 10.

[1] 1080p Video Battery Life Rundown projection (in hours) for Tears of Steel*, 1080p, 23.976fps, H.264, 3000kbps, rev1.mp4 >/=200nit, DPST=ON, no dimming, balanced, Wi-Fi connected. 11.6" 19x10 eDP panel and 36 WHr battery assumed.

SPEC CPU2006 estimates based on measurements on reference boards.

INSIDE SKYLAKE



INSIDE SKYLAKE: OUTLINE

- High level view
- Cache and Memory subsystem
- Power management and optimizations
- The core
- Processor Graphics
- Media Engine

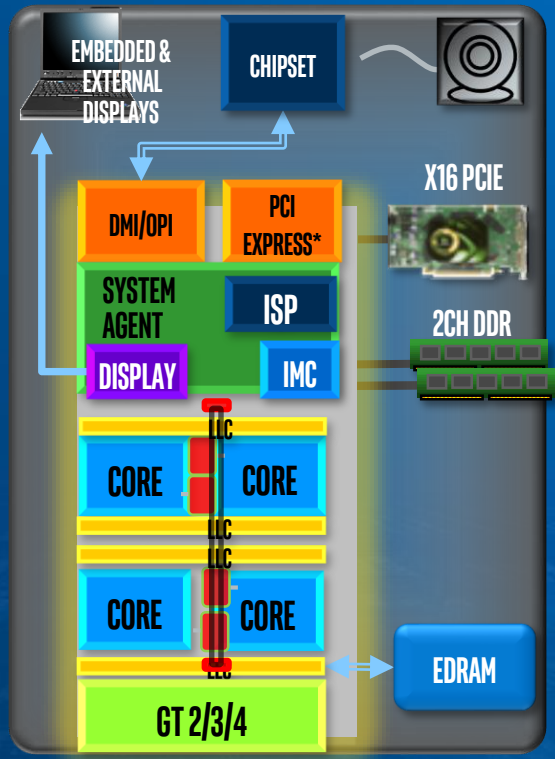
INSIDE SKYLAKE: ORIENTATION MAP

- **High level view**
- **Cache and Memory subsystem**
- Power management and optimizations
- The core
- Processor Graphics
- Media Engine

INTEL'S SKYLAKE MICROARCHITECTURE

HIGH LEVEL VIEW

- Increased chipset I/O throughput, Tablet I/Os, Audio DSP Upgrade, Sensor Hub
- Higher resolution display
- Bigger/wider core, better instruction per clock, improved power efficiency
- Enhanced ring/LLC for improved throughput

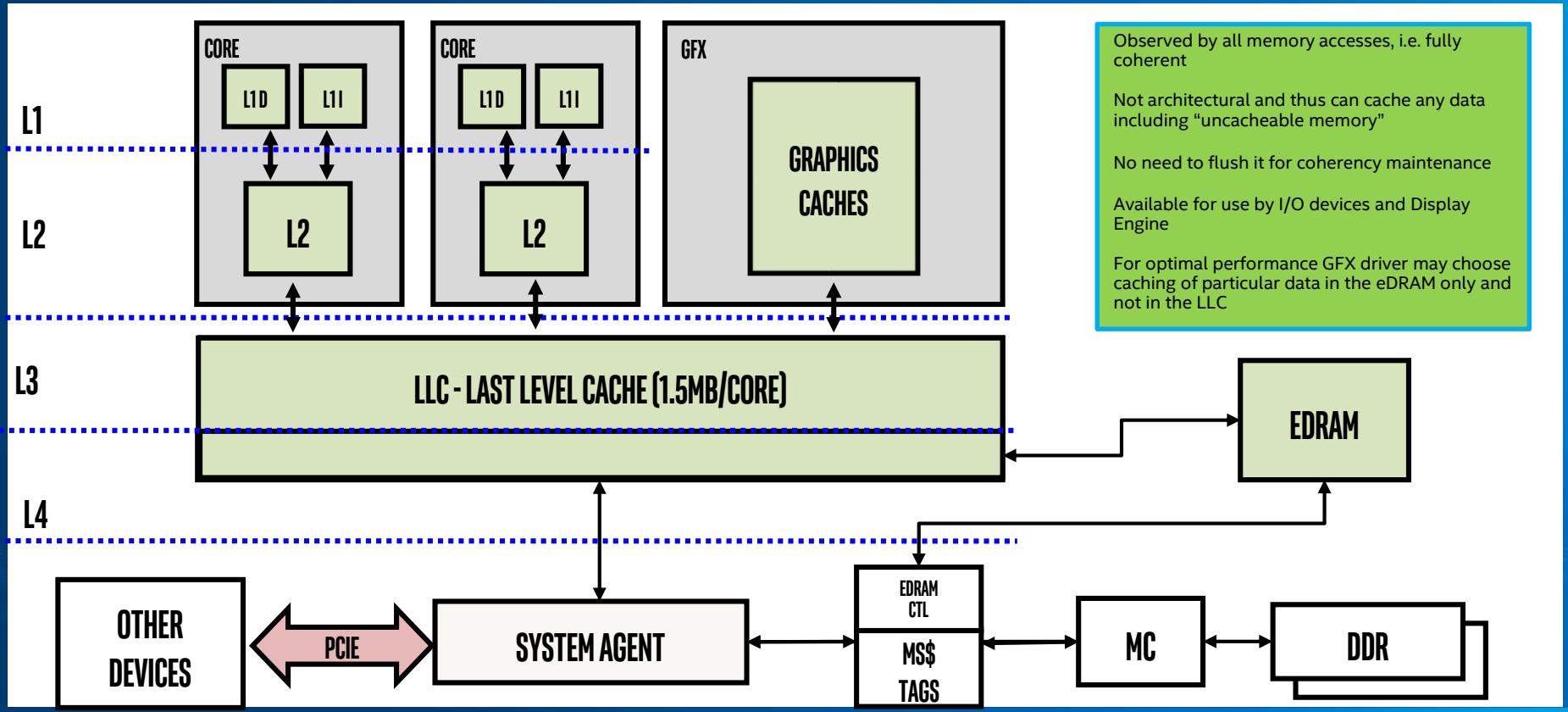


- Integrated camera ISP
- Extended overclocking capabilities
- Faster DDR Memory
- Advanced Processor Graphics GT3 + eDRAM, GT4 + eDRAM; OpenCL™ 2.0 API, DirectX® 12, OpenGL* 4.4

CACHE AND MEMORY SOLUTIONS

- Improved coherent fabric efficiency
  - Double throughput of the last level cache (LLC) miss handling
  - About 50% improvement in bandwidth per watt
- New eDRAM cache architecture, higher bandwidth, more products with eDRAM
- Support of faster DDR memory (up to DDR4 2400)
- Memory Encryption Engine for Intel® Software Guard Extension (Intel® SGX) and other security-critical data
- Improved memory QoS for high resolution displays and integrated image signal processor (ISP)
  - Higher concurrent bandwidth

EDRAM AS MEMORY SIDE CACHE



Observed by all memory accesses, i.e. fully coherent

Not architectural and thus can cache any data including “uncacheable memory”

No need to flush it for coherency maintenance

Available for use by I/O devices and Display Engine

For optimal performance GFX driver may choose caching of particular data in the eDRAM only and not in the LLC

INSIDE SKYLAKE: ORIENTATION MAP

- High level view
- Cache and Memory subsystem
- **Power management and optimizations**
- The core
- Processor Graphics
- Media Engine

POWER MANAGEMENT
INTEL® SPEED SHIFT TECHNOLOGY

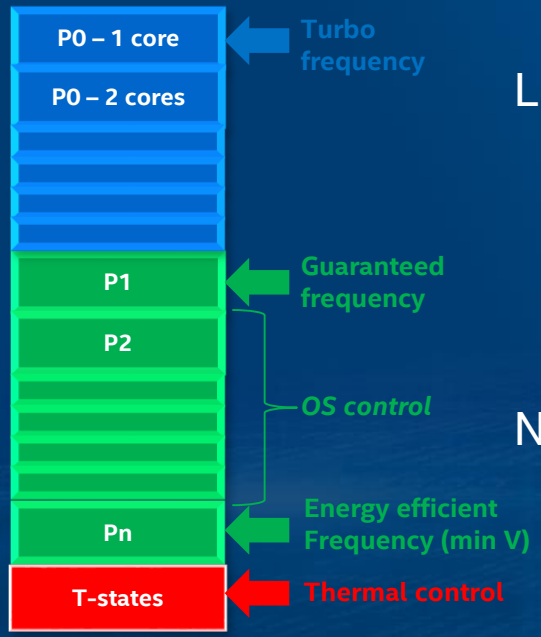
A New revolutionary approach to Power/performance management

- Hardware P-state control - A new power architecture and interface
- Fully autonomous hardware controlled P-state selection
- Operating system set minimum QoS, maximum frequency and power/performance hint
- Better observation of micro architectural behavior of workloads
- Fine grain run time control

Higher performance and responsiveness at power constrained form factors

ARCHITECTURAL INTERFACE

LEGACY INTERFACE



DVFS – Intel SpeedStep® Technology

- $P \sim V^2 \cdot f \cdot C_{dynn} + leakage(V) \sim f^3$

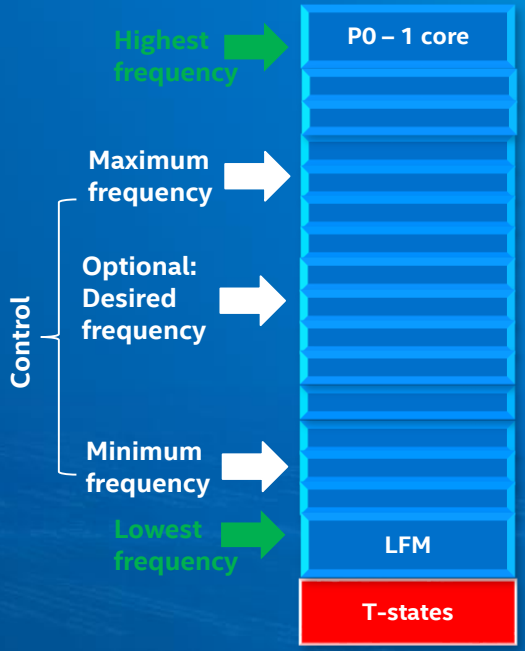
Legacy: OS controls P-state

- P1-Pn enumerated via ACPI tables
- Explicit P-state selection to P1
- Autonomous control in turbo range

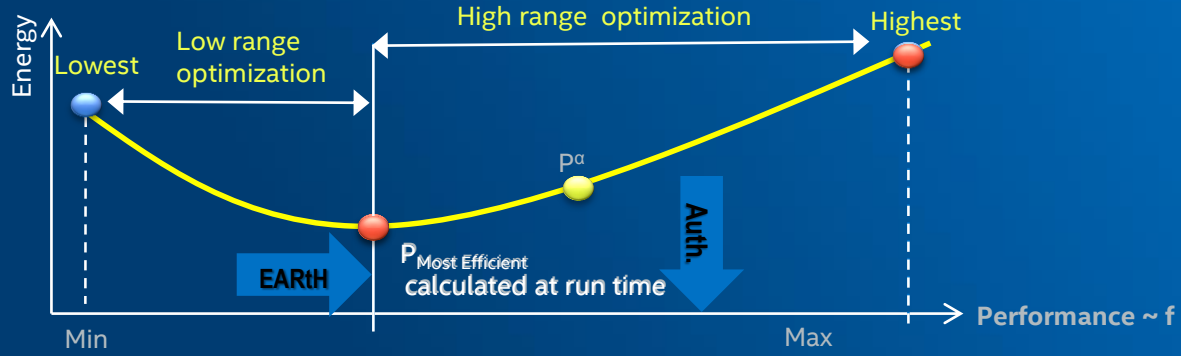
New: guided autonomous control

- OS provides min, max and preference
- Demand based HW control

HW P-state



AUTONOMOUS ALGORITHMS – ENERGY MANAGEMENT



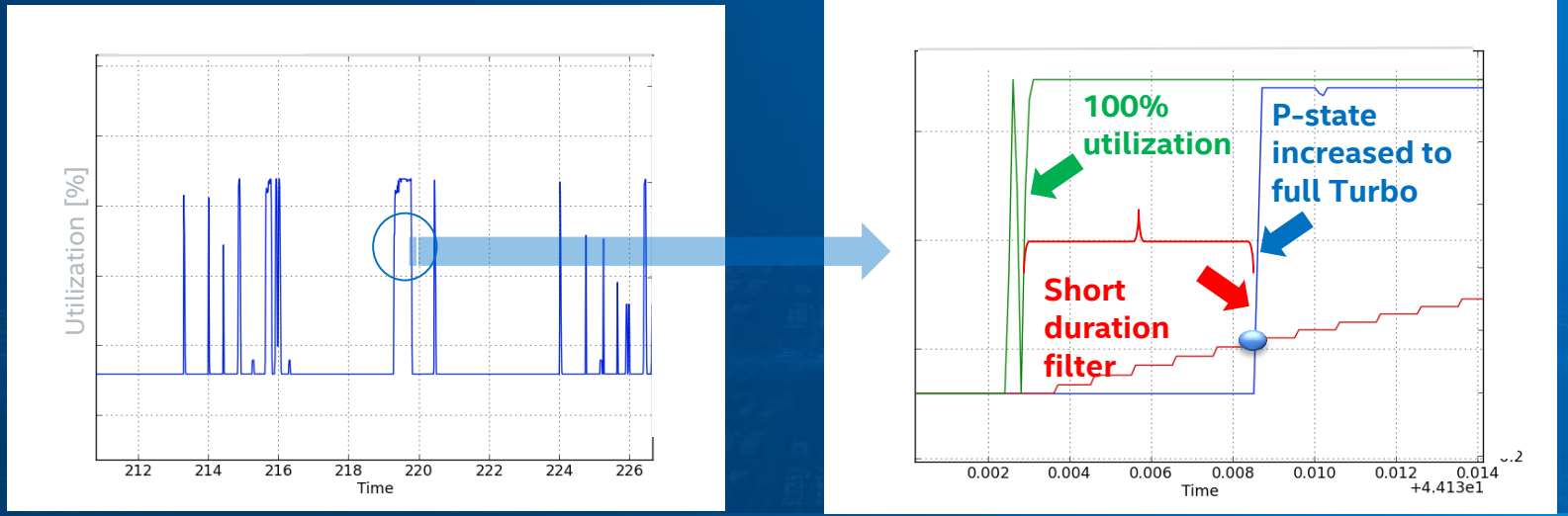
- Most energy efficient frequency (P_e) is calculated at run time (EARtH algorithm[1])
- No benefit in running lower than P_e and lose energy → frequency clamped at P_e
  - Unless critical power saving is needed
  - Only if possible to enter package sleep state (**Consumer Producer**)
- Performance = energy: Preference (α) allows limiting energy/performance
  - Semantics: frequency that meets ΔPower/Δperformance ≤α

[1] *E. Rotem, R. Ginosar, U. C. Weiser and A. Mendelson, "Energy Aware Race to Halt: A Down to EARtH Approach for Platform Energy Management," IEEE Computer Architecture Letters, vol. 99*

RESPONSIVENESS

Fast burst response while performing interactive work

- Filter out short interrupts and repeated work such as Video playback
- Filter cyclic workloads e.g. video playback



POWER OPTIMIZATIONS IN THE SKL CORE

Dynamic consumption based resource configuration

- Power Gating of Intel® AVX2 (Intel® Advanced Vector Extensions 2) hardware when it is not used
- Downscaling of underused resources

Improved scenario based power (e.g. media playback) for great mobile experience

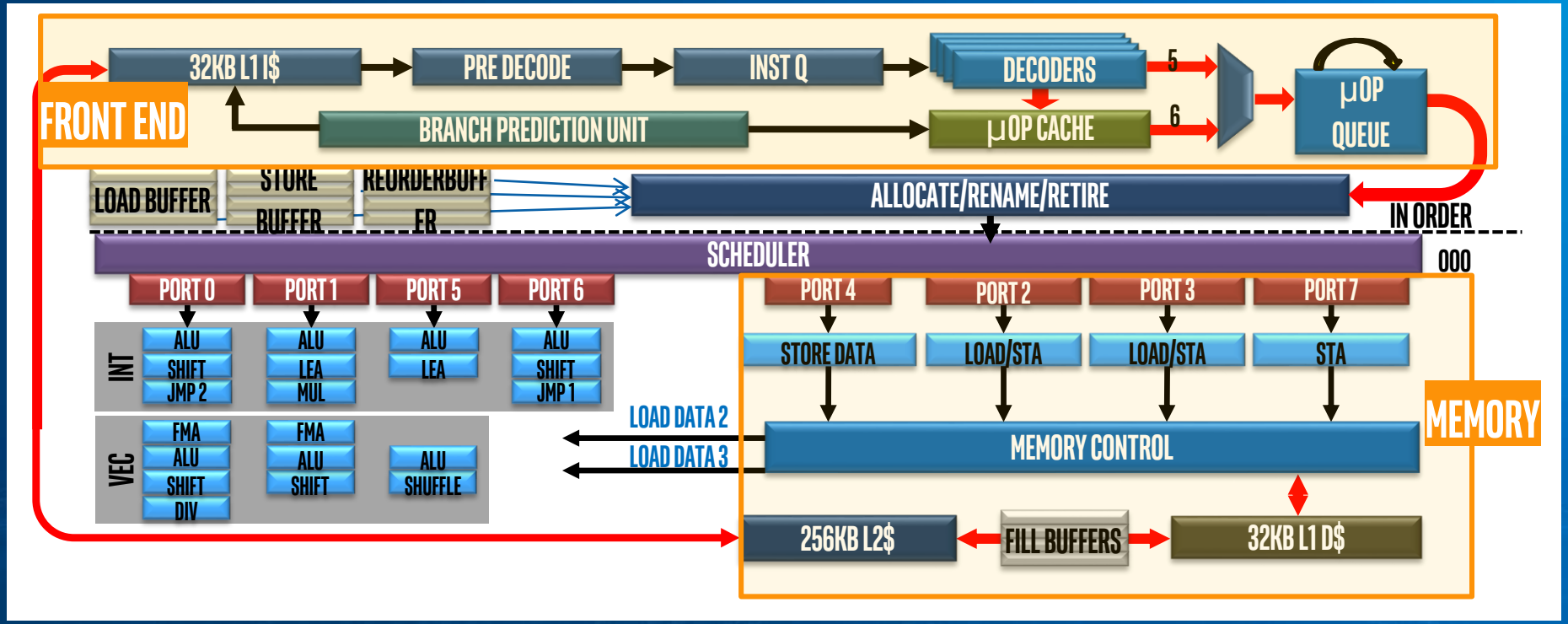
- Idle power reduction
- C1 state power reduction: improved dynamic capacitance (C_{dyn})

Better performance/Watt for the core, including focus on power at low utilization

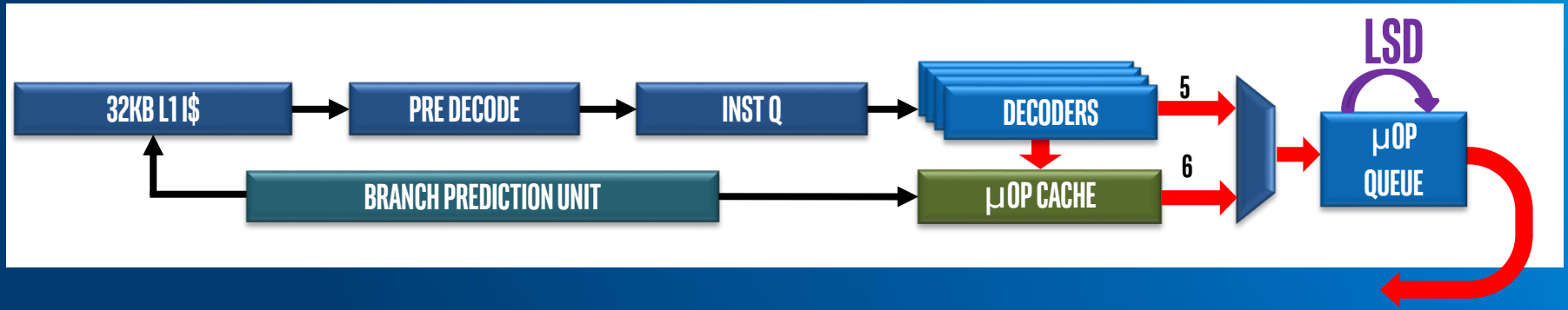
INSIDE SKYLAKE: ORIENTATION MAP

- High level view
- Cache and Memory subsystem
- Power management and optimizations
- **The core**
- Processor Graphics
- Media Engine

SKYLAKE CORE



SKYLAKE CORE: FRONT-END



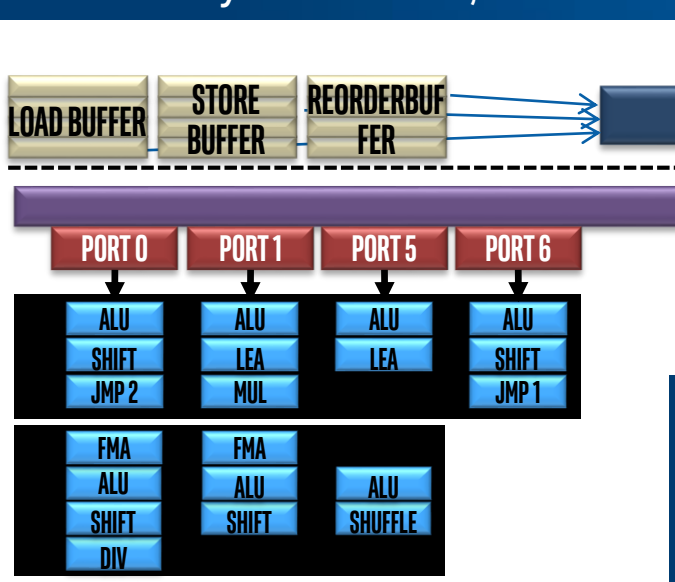
IMPROVED FRONT-END

- Increased bandwidth of Instruction Decoders and µop-cache
- Higher capacity, improved Branch Predictor
- Reduced penalty for wrong direct jump target prediction
- Faster instruction prefetch
- Increased capacity of the µop queue / Loop Stream Detector

SKYLAKE CORE: OUT-OF-ORDER EXECUTION

Deeper Out-of-Order buffers extract more instruction parallelism

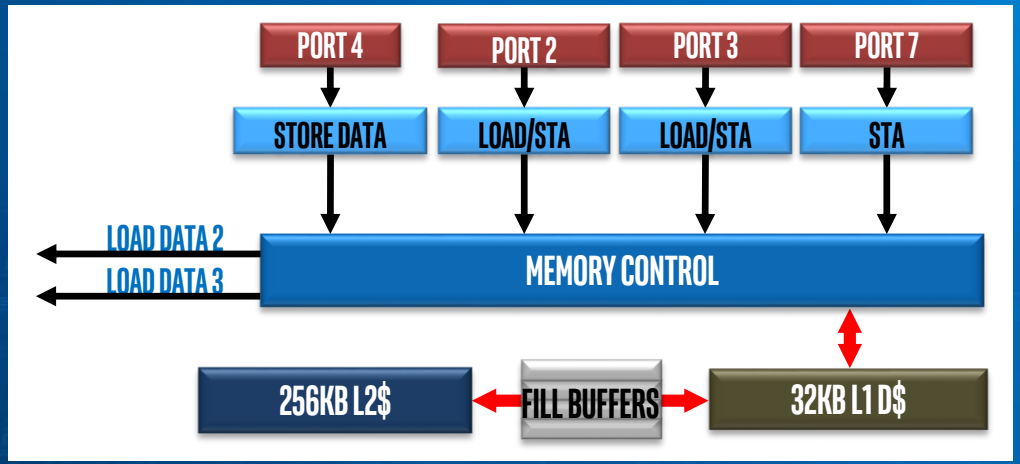
- 97 entry scheduler, 224 entry Reorder Buffer



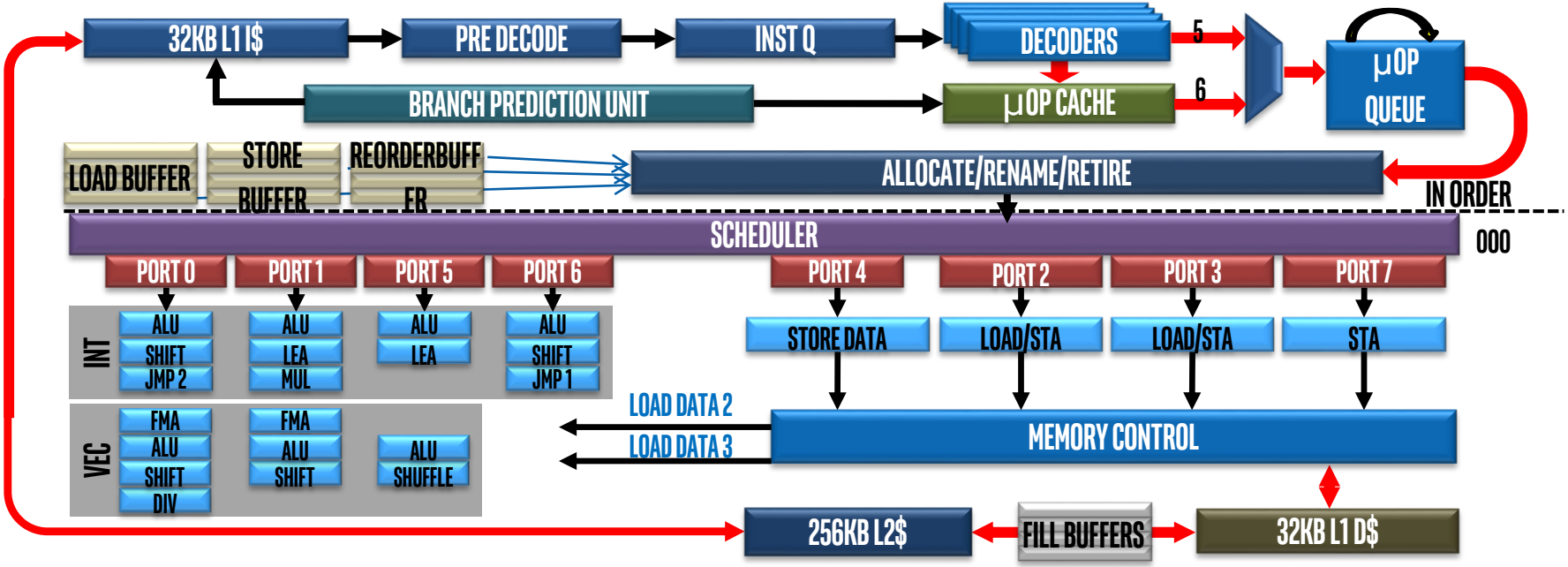
- Improved throughput and latency for divide and SQRT
- Balanced throughput and latency of Floating point ADD, MUL and FMA
- Significantly reduced latency for AES instructions

SKYLAKE CORE: MEMORY SUB-SYSTEM

- Deeper load and store buffers, increasing loads and stores throughput
- Reduced penalty for page-split loads
- Reduced Store-to-load forwarding latency
- Stores that miss in the L1$ generate L2$ requests earlier
- Higher BW from L2$ and L3$
- Higher write BW to L3$



SKYLAKE CORE



NEW SECURITY TECHNOLOGIES

Intel® Software Guard Extensions (Intel® SGX)

- A new infrastructure for trusted applications
- Instructions and flows to create, launch and operate “Enclaves” and protect them from malware and privileged software
- Provides for confidentiality, integrity, replay protection and attestation

Intel® Memory Protection Extensions (Intel® MPX)

- Memory buffer boundary testing prior to memory accesses, to ensure the memory access falls within the legitimate bounds of the object
- Helps improve SW robustness and prevent certain malware

NEW INSTRUCTIONS AND ENHANCEMENTS

XSAVE/RESTORE IMPROVEMENTS:

- Supports new Architectural state in SKL
- New instruction XSAVEC stores state in a compact layout

CLFLUSHOPT: A WEAKLY ORDERED VARIANT OF CLFLUSH

PERFORMANCE MONITORING UNIT (PMU)

SKYLAKE PMU ENABLES NEW USAGES:

Enhanced Top-Down Analysis Method [1]
- A structured approach for microarchitectural performance bottleneck detection
- Better quality and coverage for underlying PerfMon events

PMU virtualization and sharing are now easier

Precise attribution of front-end events to original code (e.g. i-cache misses)

Higher resolution event-based sampling

New class of usages with accurate Basic-Block timing in Last Branch Records
  – E.g.: Time function calls, Min/Max/Avg time per basic block

COMPARISON OF SKYLAKE'S PMU TO PREDECESSOR

	FEATURE	PREVIOUS	SKYLAKE
PERFMON	ARCH PERFMON VERSION	3	4
	EVENTS QUANTITY (COVERAGE)		RICHER
	EVENTS QUALITY		BETTER
	TOP-DOWN ANALYSIS	BASIC SMT-OFF	ACCURATE SMT ON
PEBS	FRONT-END EVENTS COVERAGE	NO	YES
	TSC INCLUDED IN RECORD	NO	YES
LBR	TIMING INFO	NO	YES
	# ENTRIES	16	32

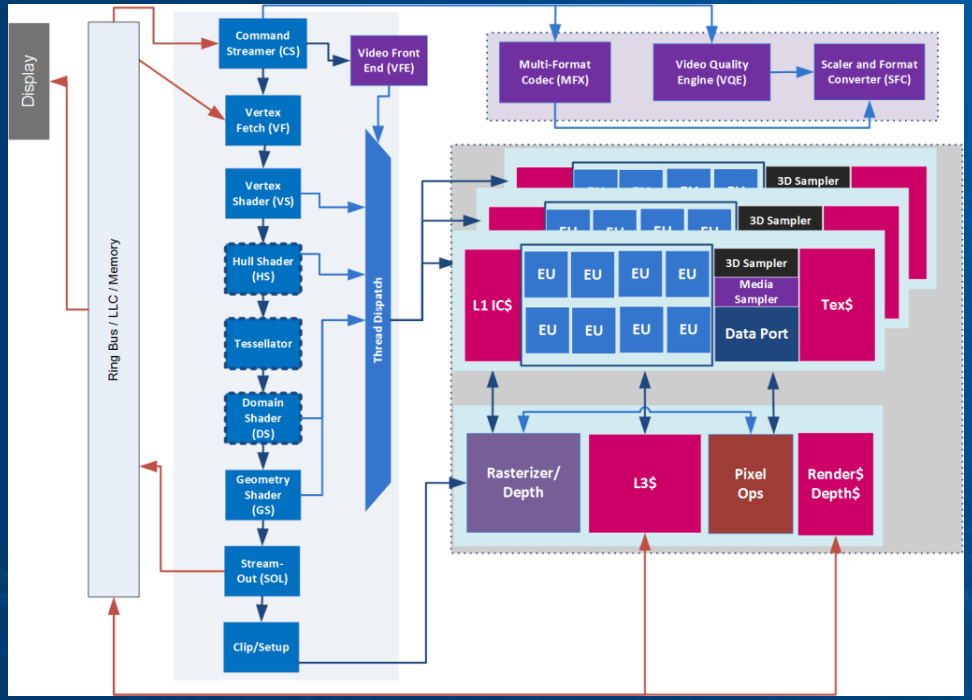
PEBS= Precise Events Based Sampling
LBR= Last Branch Records

[1] A. Yasin, "A Top-Down Method for Performance Analysis and Counters Architecture", ISPASS 2014

INSIDE SKYLAKE: ORIENTATION MAP

- High level view
- Cache and Memory subsystem
- Power management and optimizations
- The core
- **Processor Graphics**
- **Media Engine**

SKYLAKE GRAPHICS

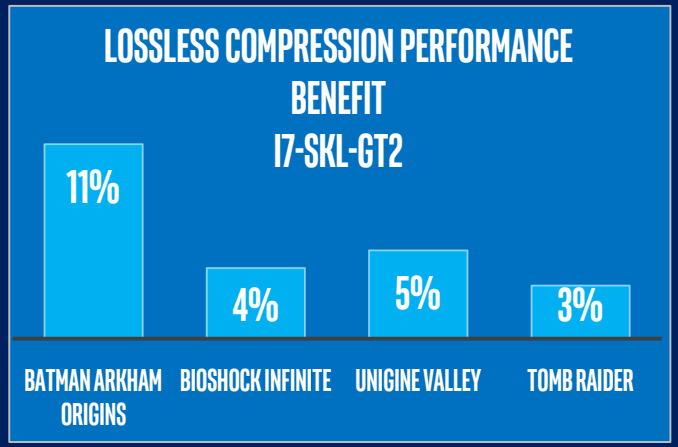


- **Scalable performance from 24-72 EUs**
  - 384-1152 SP FLOPS/clock
  - Geometry throughput increases
  - Pixel fill rate increase by 1.33-2x/clock
  - Lossless render compression
- **New 3D Application Features**
  - Conservative Rasterization
  - Bindless Textures
  - Render Target read
- **API Support**
  - DirectX* 12, Vulkan, Metal, OpenGL
  - OpenCL™ 2.0, extends CPU/GPU Programmability

SKYLAKE GRAPHICS
MICROARCHITECTURE IMPROVEMENTS

LOSSLESS RENDER COMPRESSION

- Texture, pixel read/write paths
- 2:1 peak compression ratio
- Saves bandwidth and memory power

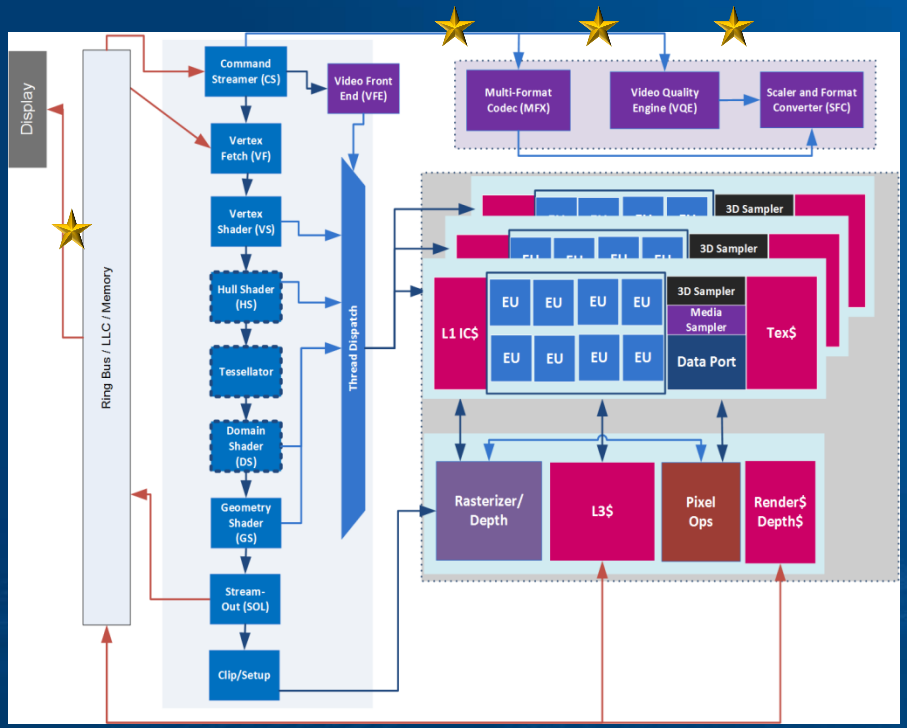


FINER GRAIN PREEMPTION

- Increases responsiveness of system
- 3D: triangle boundaries
- Compute: Thread-group → Mid-thread

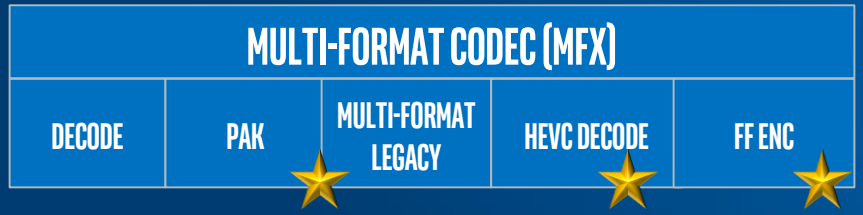
	THREAD GROUP		MID THREAD	
	U-SERIES	Y-SERIES	U-SERIES	Y-SERIES
ADOBE PHOTOSHOP*	4-6MS	17-22MS	300US	800US
SAMPLE APP1	200-500MS	200-500MS	300US	280-320US
SAMPLE APP2	17MS	24MS	240US	200-430US

SKYLAKE MEDIA SYSTEM-LEVEL IMPROVEMENTS



- **Introducing Skylake with Gen9 Processor Graphics, built on 14nm**
- **Improved power management to optimize media usages**
  - Ability to run only fixed function (FF) media in the Unslice without powering on the Slices for key workloads
  - Ability to run selective slices instead of full slices
  - Ability to use selective pair of EUs in the subslice
  - Inter-media engine memory compression to reduce bandwidth
- **New Intel® Quick Sync Video mode**
  - Now support Processor Graphics (PG) and Fixed Function (FF) modes
- **Broad Enabling of Applications**
  - Support for DirectX* 11.2, DirectX-next and OpenCL™ 2.0, extends CPU/GPU Programmability

SKYLAKE MEDIA ENGINE IMPROVEMENTS



MFX

- HEVC 8b decode/encode for real-time 4Kp60 usages
- VP8 and JPEG/MJPEG decode/encode
- Fixed function low power low latency AVC encoder latency real-time usages

VQE

- 16b processing path
- Spatial 5x5 denoise filter
- Local Adaptive Contrast Enhancement (LACE)
- Camera processing features
- Improved camera/image RAW processing capability

NEW

SFC

- New in-line scaling & format conversion engine
- Includes IEF sharpening filter

SUMMARY

- Skylake delivers record levels of performance and battery life in many personal computing use cases and form factors
- Intel® Speed Shift Technology provides higher performance, responsiveness and efficiency at power constrained form factors
- Skylake Processor Graphics delivers scalable performance, >1TFLOPS compute, enhanced low power media engines, flexible power management, and end-to-end 4K experience
- Skylake family of products allows developers to:
  - Choose from wide range of platform capabilities
  - Innovate with products for wide range of thermal envelopes and I/O solutions
  - Optimize the system performance using the advanced PMU capabilities
- Skylake introduces Intel® SGX: a revolutionary game changer to trusted application security in the main stream SW environment


intel®
experience
what's inside™